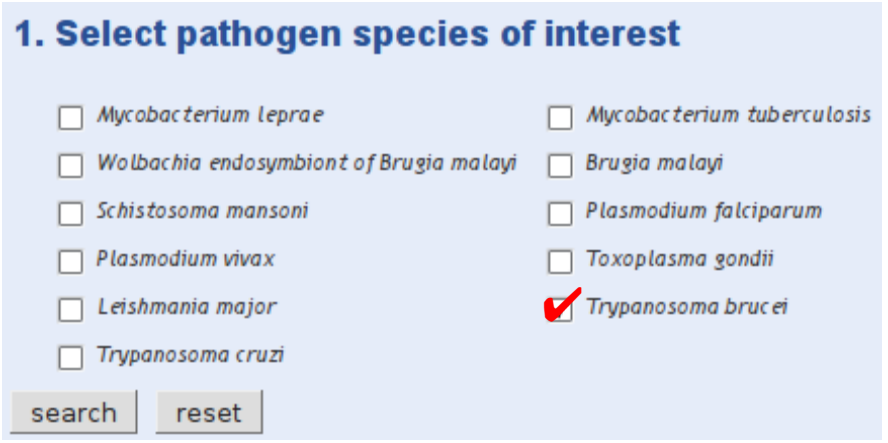
Supplementary Figure S1

This figure demonstrates the use of the TDR Targets Database to produce a ranked list of putative drug targets for *Trypanosoma brucei*. In this example, multiple searches on the genome of *T. brucei* are performed in each of the following categories: Name/Annotation, Features, Structures, Phylogenetic distribution, Essentiality, Druggability, Validation Data, and Bibliographic references. For enhanced readability we have put each of these searches in a separate page, starting with the next page.

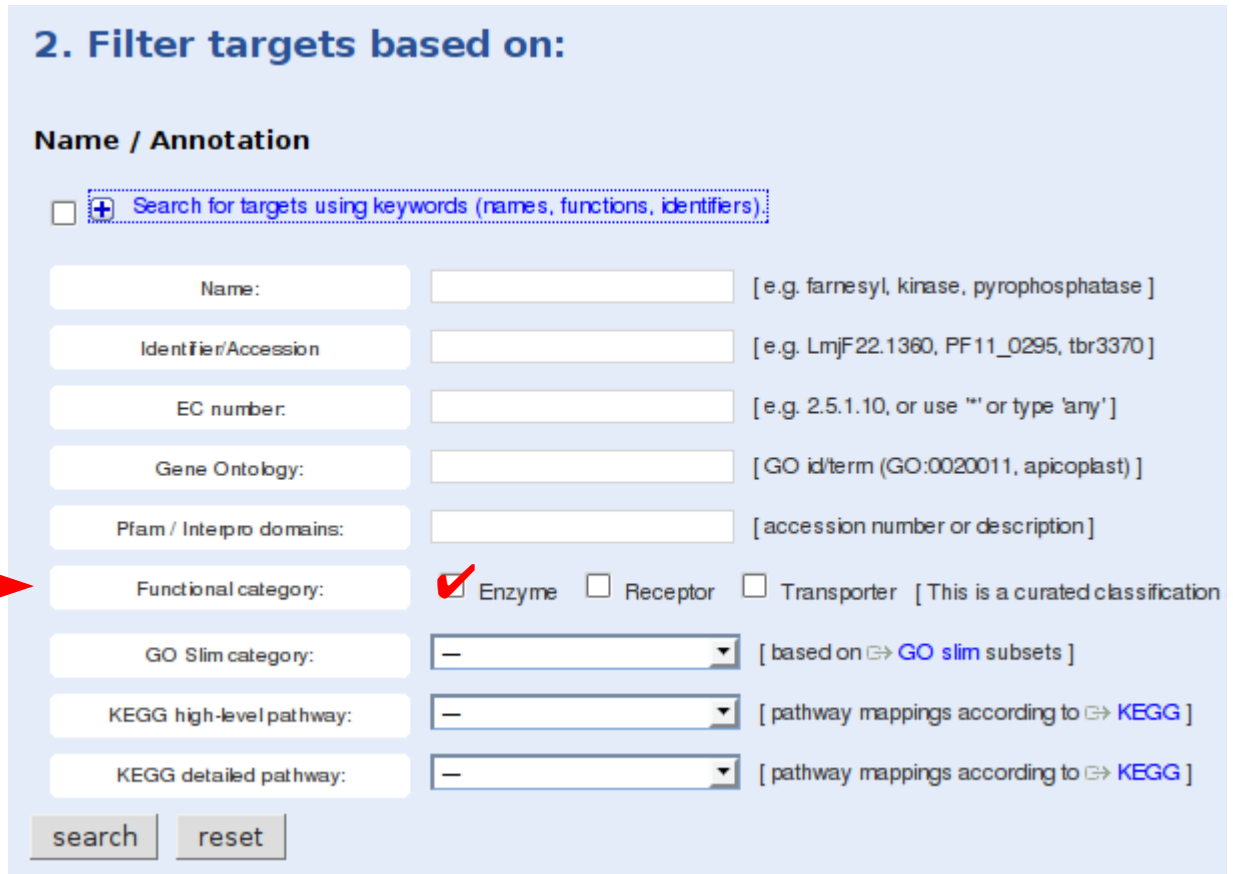
When performing these searches online, each of them will be automatically saved in the user's query history. On the history page, the queries can then be individually weighted, and then combined to obtain a ranked-order of the union of all queries.

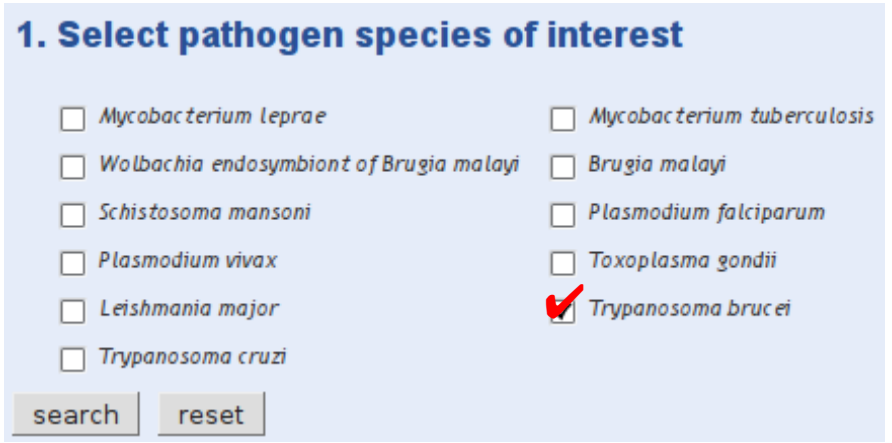
Other usage examples, in the form of slideshow tutorials, are available online at:
http://tdrtargets.org/tutorials, or at
http://slideshare.net/tdrtargets



Search for *T. brucei* enzymes.

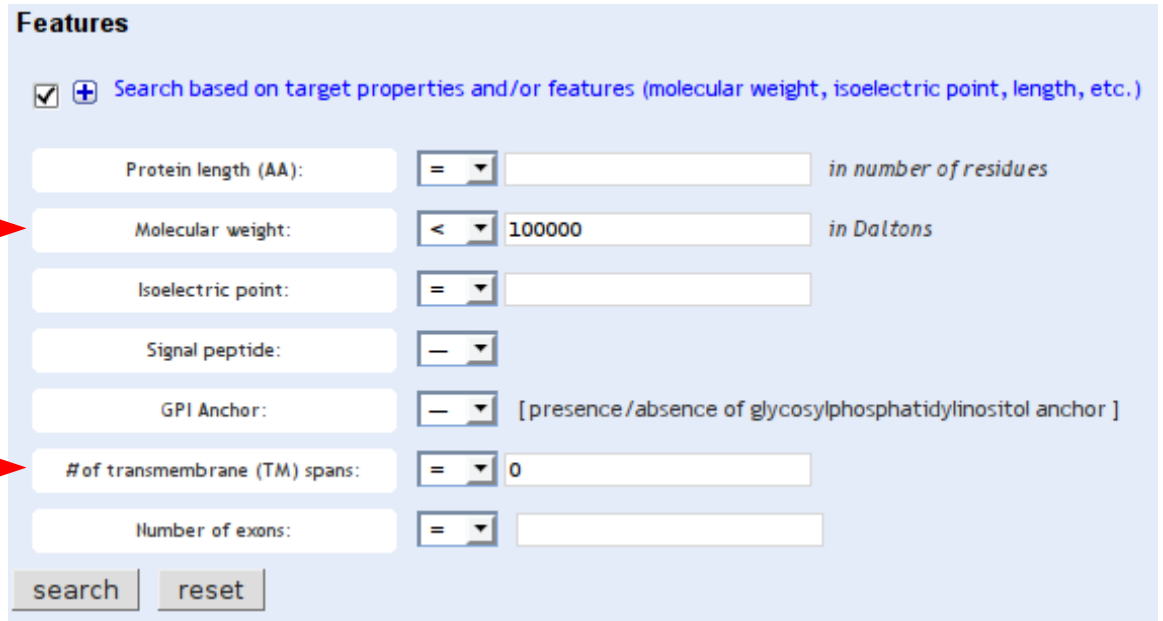
1. Select species of interest
2. Specify criterion under **Name/Annotation**
3. (Optional) Name your query
4. Run





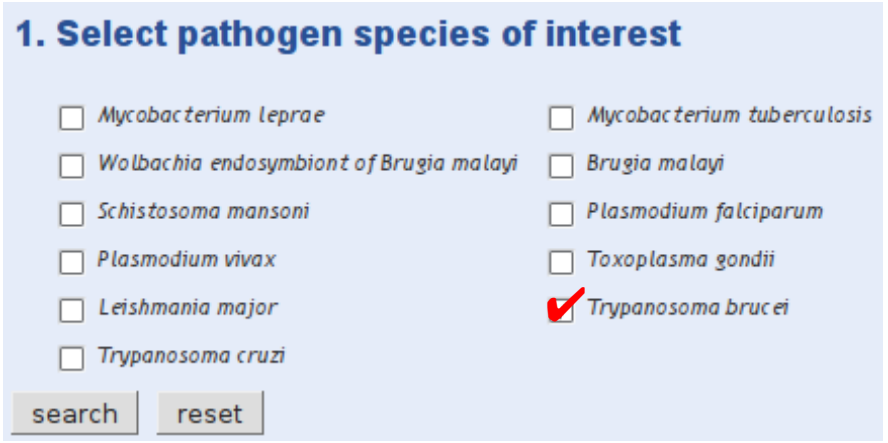
Filter *T. brucei* genes that are likely to be expressed as soluble proteins in recombinant form: low molecular weight and no transmembrane domains.

Two separate queries are run to express this criterion, to allow each query to be weighted independently.



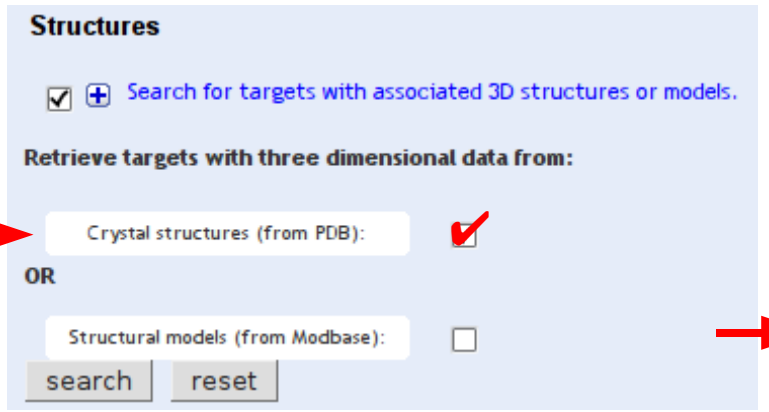
Low MW query
1. Select species of interest
2. Specify criterion under **Features**
3. (Optional) Name your query
4. Run

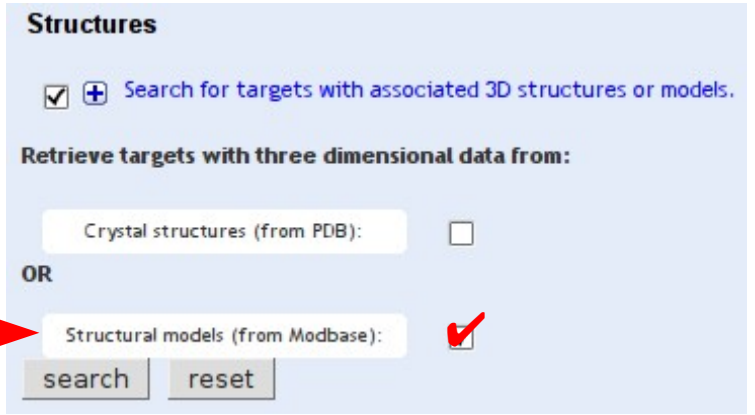
No TM domains query
1. Select species of interest
2. Specify criterion under **Features**
3. Name the query
4. Run

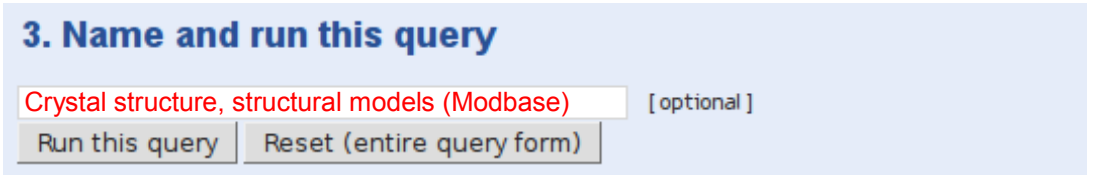


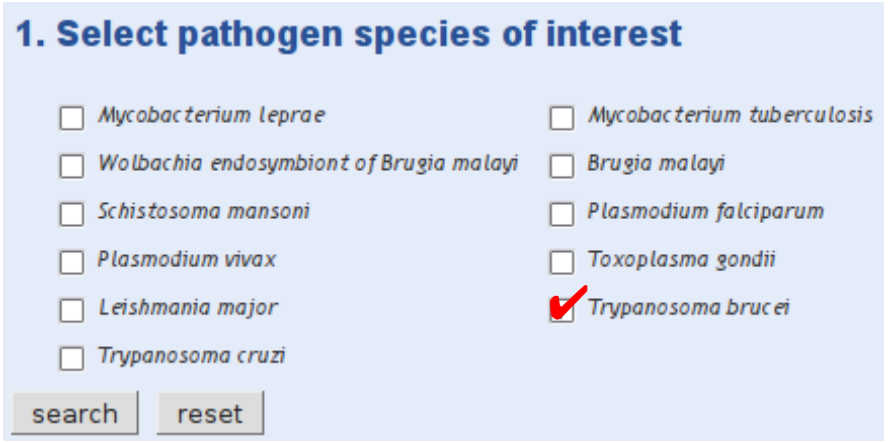
Filter *T. brucei* genes based on the available 3D structural information available: crystal structures or structural models.

Two separate queries are run to express this criterion, to allow each query to be weighted independently.



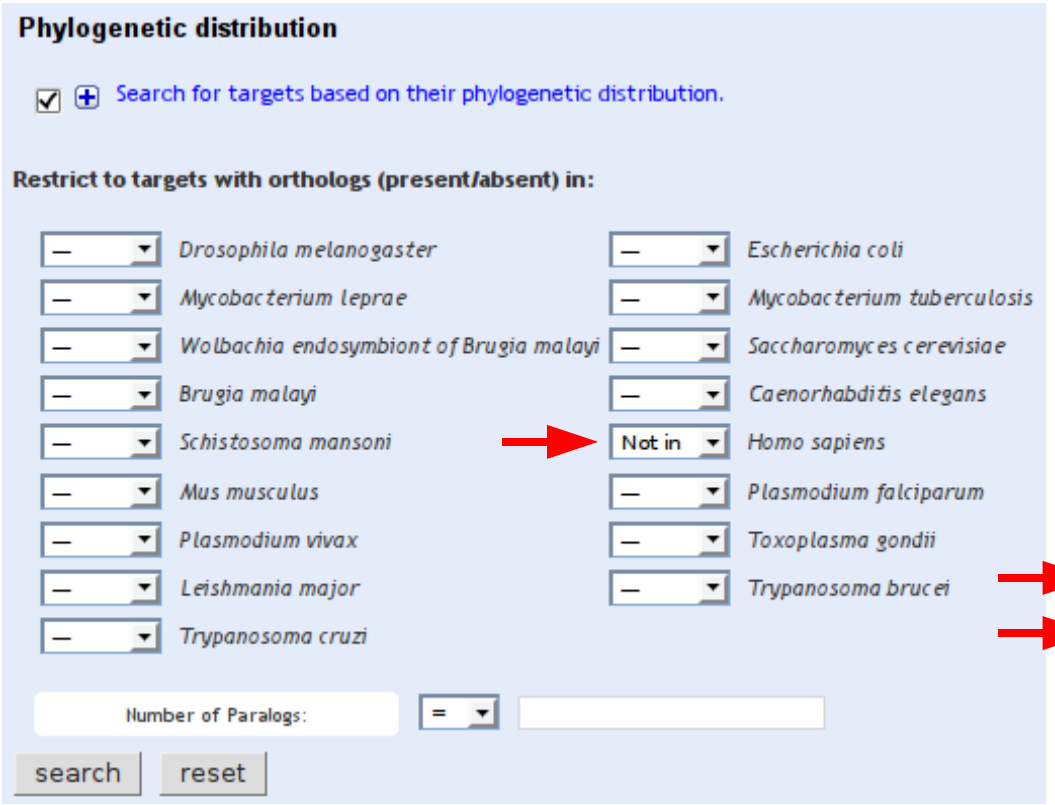


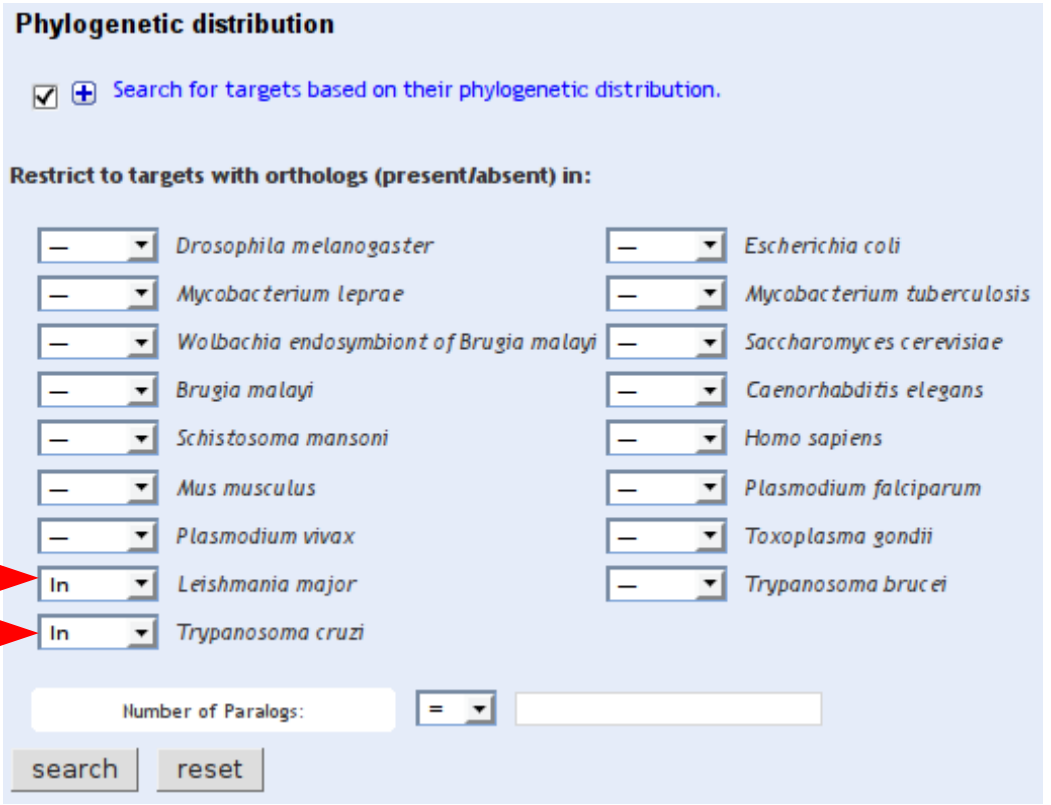


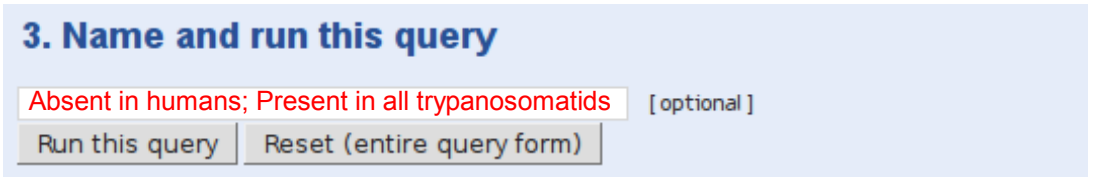


Search for *T. brucei* genes based on their phylogenetic distribution: absent in humans and present in other trypanosomatids.

Two separate queries are run to express this criterion, to allow each query to be weighted independently.







1. Select pathogen species of interest

- [ ] *Mycobacterium leprae*
- [ ] *Mycobacterium tuberculosis*
- [ ] *Wolbachia endosymbiont of Brugia malayi*
- [ ] *Brugia malayi*
- [ ] *Schistosoma mansoni*
- [ ] *Plasmodium falciparum*
- [ ] *Plasmodium vivax*
- [ ] *Toxoplasma gondii*
- [ ] *Leishmania major*
- [x] *Trypanosoma brucei*
- [ ] *Trypanosoma cruzi*

search | reset

Search for *T. brucei* genes with at least one essential ortholog in a model organism.

Essentiality

[x] Search for targets that are essential/inviable.

Retrieve targets for which genome-wide information about their essentiality is available.
If genome-wide information for an organism is not available, you can evaluate the essentiality of
from the options below. Also note that essential genes for your organism of interest might show
option further down).

[x] **Any** evidence of essentiality in **any** species

Or

Select the species and the type of 'essential' phenotype from the options below: note that if you
the UNION (boolean OR) of the selection.

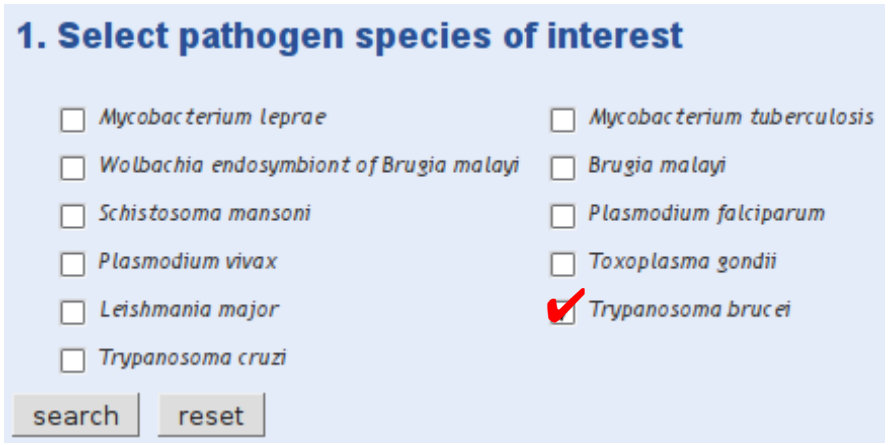
C. elegans	—
E. coli	—
M. tuberculosis	—
S. cerevisiae	—

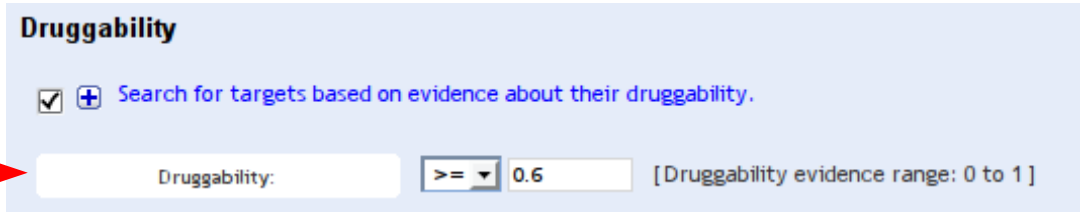
search | reset

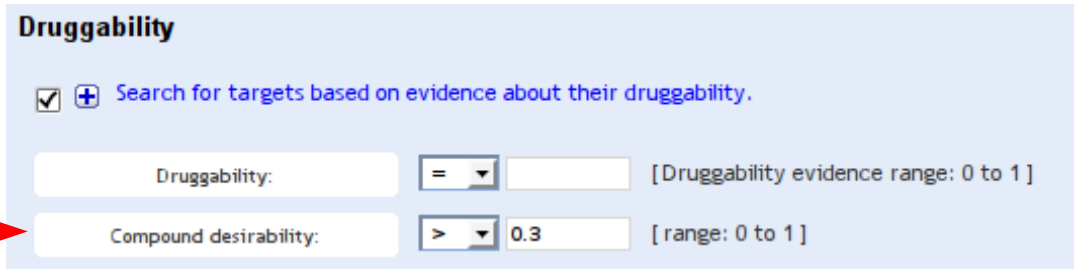
3. Name and run this query

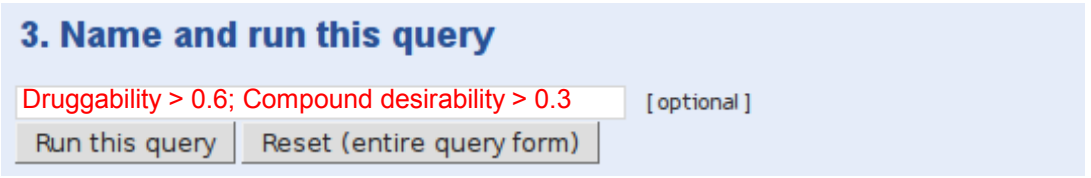
Essential in at least one model organism [optional]

Run this query | Reset (entire query form)



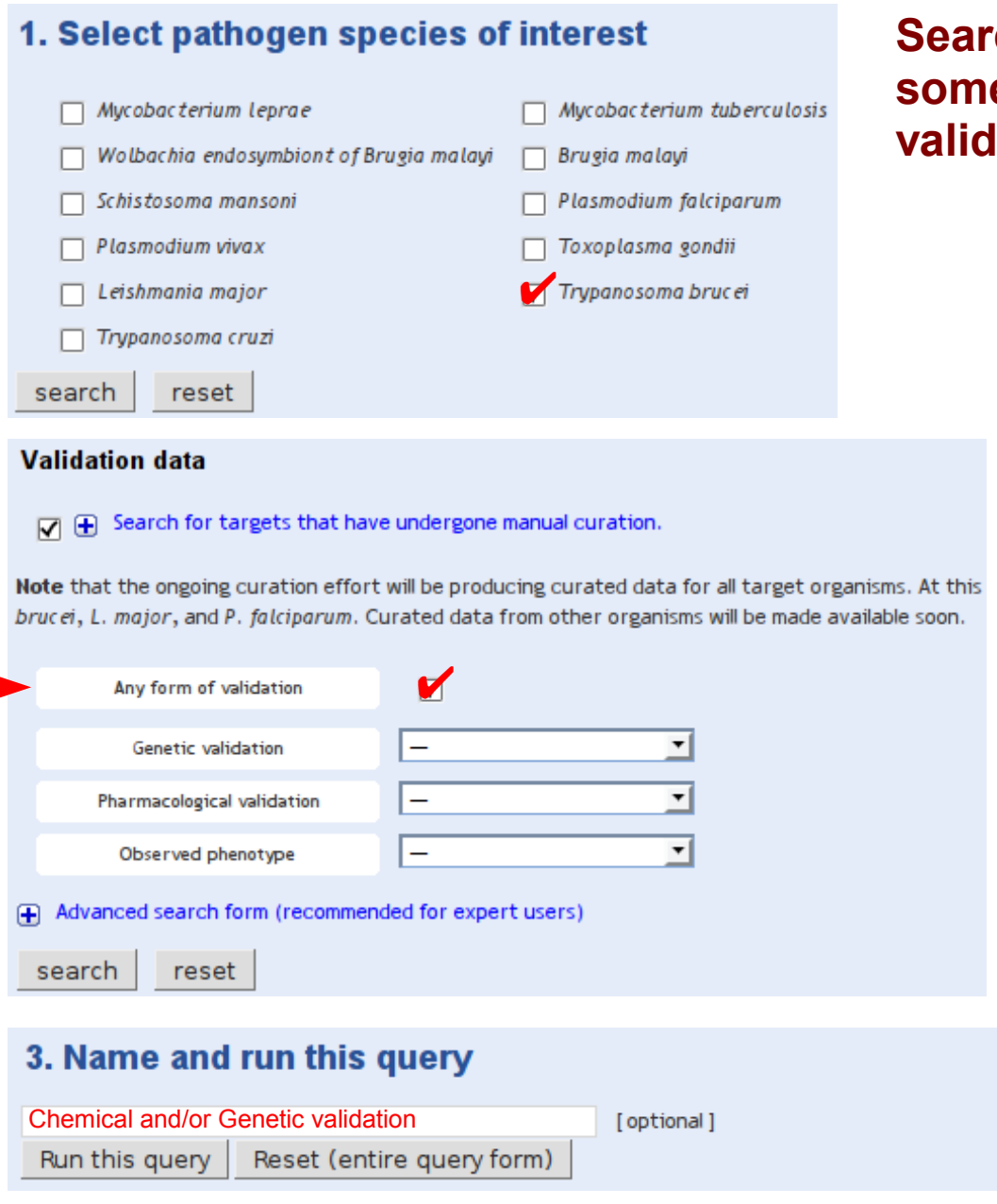




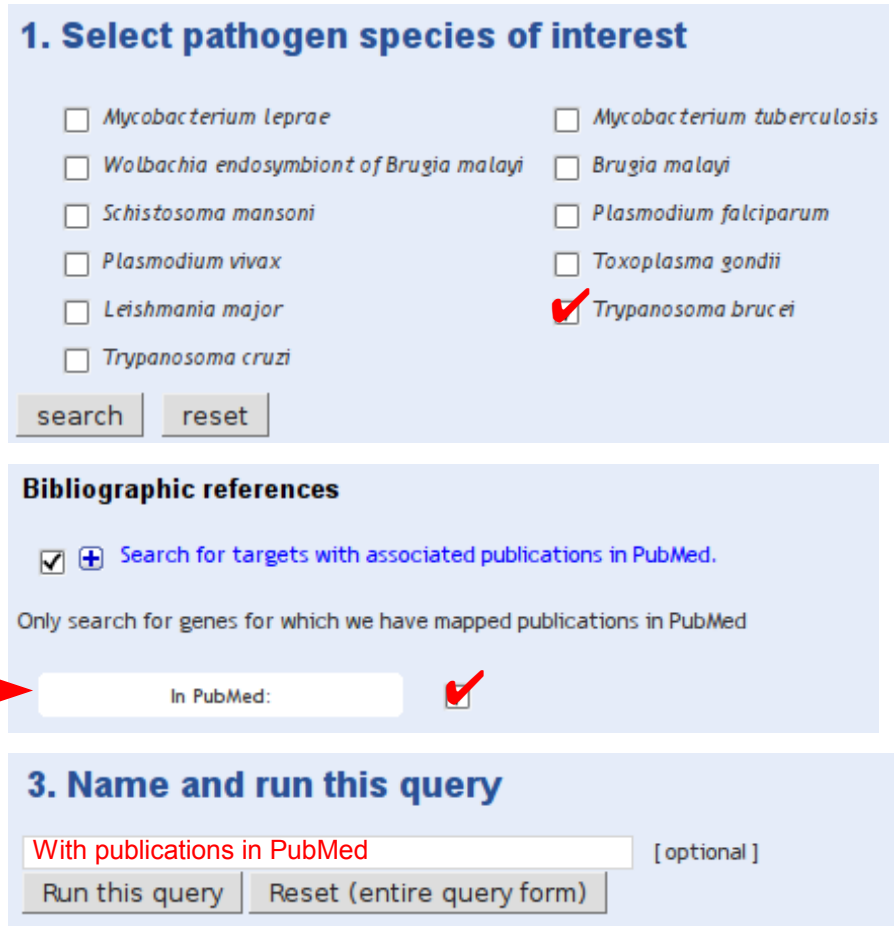


Search for *T. brucei* genes with some precedence for druggability.

Two separate queries are run to express this criterion, to allow each query to be weighted independently.



Search for *T. brucei* genes for which there is some curated information about their validation credentials.



Search for *T. brucei* genes for which there are bibliographic references.

Note that this query also relies on manual association of genes with references.

The individual weights assigned to each query reflect the preferences of the user. In this example, enzymes are weighted most highly (100) because they often have druggable active sites and are assayable. Features such as having relatively low mass (<100 kDa) and having no transmembrane domains (no TM) may make the protein easier to express in recombinant systems, so each is given some weight (20 each). Having a experimentally determined crystal structure (50) is weighted more highly than having a model available (30) because structure-based drug design is more likely to be successful with a crystal structure than a model. Under Phylogenetic Distribution, genes that are found in all trypanosomatids but not in humans are weighted 25 each, as these may be more likely to be broad-spectrum parasite-specific targets. If orthologs of the targets are essential in any of the model species for which we have collected data, these targets are given moderate weights (40) recognizing an increased likelihood that these may be lethal targets when inhibited. If orthologs have been found to be more "druggable" (target precedence) (35) or have desirable compounds associated as ligands (35), they are given moderate weights. If targets have been manually curated and found to have been validated as a drug target in some way (either chemically or genetically) in the organism, they are also weighted (50). Finally, if there are publications available in PubMed, some work may already be done on the target, so this is weighted too (35).

Combine and weight previous queries in the website history page.

Criterion	Weight
Enzymes	100
Low molecular weight	20
No transmembrane domains	20
Crystal structure	50
Structural model (Modbase)	30
Present in all trypanosomatids	25
Absent in humans	25
Essential in at least one model organism	40
Druggability index > 0.6	35
Compound desirability index > 0.3	35
Chemical and/or genetic validation	50
With publications in PubMed	35
Maximum possible cumulative score	**465**

A union of these weighted queries maybe generated on the “History” page, providing a list of all *Trypanosoma brucei* genes, ranked according to their priority as a drug target using the above values. The highest scoring target in this particular exercise is farnesyl pyrophosphate synthase, a protein that additional experimental work suggests is a promising drug target. Clicking on the name of the target (Tb927.7.3360) leads to a target-specific page showing that as of July 2008, this target has been genetically validated by RNAi to demonstrate a growth deficit in the bloodstream (mammalian stage), that it is a 42kDa enzyme, that its orthologues in C. elegans and S. cerevisiae are essential, that it has a druggability score of 0.8 (0-1, 1 is optimal druggability) and a compound desirability score of 0.3 (in a range of 0-1) based on the interactions of 97 inhibitors with orthologues, that literature links 12 interacting chemical compounds with this enzyme in *T. brucei*, that there are 2 structures for this enzyme in PDB as well as a ModBase model, that we have curated 12 bibliographic references for the enzyme, and that genetic and chemical validation experiments have been published on this enzyme in *T. brucei*. All of the listed genes can be similarly examined in depth by clicking on their names. Finally, lists may be posted for the benefit of other users, and can be exported as a tab-delimited file for further manipulation in spreadsheet form.

Name	Product	Weight
Tb927.7.3360	farnesyl pyrophosphate synthase	405
Tb927.1.700	phosphoglycerate kinase	370
Tb10.70.4740	enolase	370
Tb10.70.1370	fructose-bisphosphate aldolase, glycosomal	370
Tb11.02.0330	UDP-galactose 4-epimerase	365
Tb927.6.2030	protein kinase, putative	365
Tb11.02.2310	prostaglandin f synthase	355
Tb927.6.560	cysteine peptidase C (CPC),CPC cysteine peptidase, Clan CA, family C1, Cathepsin ...	355
Tb10.61.2550	N-myristoyl transferase, putative	355
Tb09.160.4090	DNA topoisomerase II	345
Tb927.3.1380	ATP synthase beta chain, mitochondrial precursor,ATP synthase F1, beta subunit	340
Tb927.8.5950	protein kinase, putative	340
Tb10.61.3140	protein kinase, putative,glycogen synthase kinase, putative	340
Tb10.70.2210	cell division related protein kinase 2, putative,CDC2-related protein kinase	340
Tb927.4.1080	V-type ATPase, A subunit, putative	340
Tb927.7.6220	protein kinase, putative	340
Tb11.02.0640	protein kinase, putative,dual-specificity protein kinase, putative	340